

Strained Silicon Structure

TECHNICAL FIELD

[0001] The present invention generally relates to the fabrication of semiconductor devices. In one aspect it relates more particularly to a strained silicon structure.

BACKGROUND

[0002] Complementary metal-oxide-semiconductor (CMOS) technology is a dominant semiconductor technology used for the manufacture of ultra-large scale integrated (ULSI) circuits today. Size reduction of the metal-oxide-semiconductor field-effect transistor (MOSFET) has provided significant improvement in the speed performance, circuit density, and cost per unit function of semiconductor chips over the past few decades. Significant challenges are faced when CMOS devices are scaled into the sub-100 nm regime. An attractive approach for additional improvement of CMOS transistor performance exploits strain-induced band-structure modification and mobility enhancement to increase the transistor drive current. Enhanced electron and hole mobilities in silicon (Si) under biaxial tensile strain can be achieved. Enhanced electron and hole mobilities improve the drive currents of N-channel and P-channel MOSFETs, respectively. In the strained silicon, electrons experience less resistance and flow up to 70 percent faster, which can lead to chips that are up to 35 percent faster without having to further shrink the size of transistors.

[0003] As shown in FIG. 1A, many designs of strained silicon layers for transistor fabrication utilize buffer layers or complex multi-layer structures on a bulk silicon substrate 20. Strained silicon substrate technology often utilize a silicon-germanium (SiGe) graded buffer layer 22 with a thickness in the order of microns. A relaxed SiGe layer 24 overlies the graded

buffer layer 22. The relaxed SiGe layer 24 has a larger natural lattice constant than that of silicon. Relaxed crystalline silicon is said to be lattice-mismatched with respect to relaxed crystalline SiGe due to the difference in their lattice constants. As a result, a thin layer of silicon 26 that is epitaxially grown on the relaxed SiGe layer 24 will be under biaxial tensile strain because the lattice of the thin layer of silicon 26 is forced to align to the lattice of the relaxed crystalline SiGe layer 24, as illustrated in FIGs. 1B and 1C. Referring again to FIG. 1A, a transistor 28 is formed in the silicon layer 26. The transistor 28 includes a source 30, a drain 32 and a gate 34. Transistors fabricated on the strained silicon layer 26 will have enhanced electrical performance. As also shown in FIG. 1A, the transistor 28 is typically bounded by an isolation region 36 (e.g., shallow-trench isolation (STI), local oxidation of silicon (LOCOS), field oxide (FOX)).

[0004] The graded SiGe buffer layer 22 introduces a lattice mismatch with the underlying silicon substrate 20, which may result in a dispersed, three-dimensional misfit dislocation network. Strain-relieving glide of threading dislocations 38 is facilitated. Dislocations formed in the graded buffer layer 22 may propagate to the wafer surface, resulting in a defect density in the order of 10^4 - 10^5 defects per cm^2 . Such a high defect density may present a significant barrier for the production of integrated circuits using such substrates. Also, the underlying strain fields of the misfit arrays result in a characteristic cross-hatch surface roughness. This surface roughness can be a significant problem as it potentially degrades channel mobility in active devices. Hence, there is a need for a way to reduce defect density for such strained silicon substrate structures.

SUMMARY OF THE INVENTION

[0005] The problems and needs outlined above may be addressed by embodiments of the present invention. In accordance with one aspect of the present invention, a semiconductor device is provided, which includes a substrate, a first epitaxial layer, a second epitaxial layer, a third epitaxial layer, a first trench, and a second trench. The first epitaxial layer is formed on the substrate. The first layer has lattice mismatch relative to the substrate. The second epitaxial layer is formed on the first layer, and the second layer has lattice mismatch relative to the first layer. The third epitaxial layer is formed on the second layer, and the third layer has lattice mismatch relative to the second layer. The first trench extends through the first layer. The second trench extends through the third layer and at least partially through the second layer. At least part of the second trench is aligned with at least part of the first trench, and the second trench is at least partially filled with an insulating material.

[0006] In accordance with another aspect of the present invention, a method of manufacturing a semiconductor device, is provided. This method includes the following steps described in this paragraph, and the order of steps may vary. A substrate is provided. A first epitaxial layer is formed on the substrate. The first layer has lattice mismatch relative to the substrate. A first trench is formed in the first layer. A second epitaxial layer is formed on the first layer. The second layer has lattice mismatch relative to the first layer. A third epitaxial layer is formed on the second layer. The third layer has lattice mismatch relative to the second layer. A second trench is formed in the third and second layers. At least part of the second trench is in alignment with at least part of the first trench.

[0007] In accordance with yet another aspect of the present invention, a method of manufacturing a semiconductor device, is provided. This method includes the following steps

described in this paragraph, and the order of steps may vary. A substrate is provided. A first epitaxial layer is formed on the substrate. The first layer has lattice mismatch relative to the substrate. A second epitaxial layer on the first layer. The second layer has lattice mismatch relative to the first layer. A first trench is formed in the second and first layers. A third epitaxial layer is formed on the second layer. The third layer has lattice mismatch relative to the second layer. A second trench is formed in the third and second layers. At least part of the second trench is in alignment with at least part of the first trench.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The following is a brief description of the drawings, which illustrate exemplary embodiments of the present invention and in which:

[0009] FIG. 1A is a cross-section view of a transistor formed on a strained silicon substrate in accordance with a prior design;

[0010] FIGs. 1B and 1C illustrate a way to form a silicon layer under biaxial tension;

[0011] FIGs. 2-7 illustrate steps of forming a first illustrative embodiment of the present invention; and

[0012] FIGs. 8-11 illustrate steps of forming a second illustrative embodiment of the present invention.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0013] Referring now to the drawings, wherein like reference numbers are used herein to designate like elements throughout the various views, illustrative embodiments of the present invention are shown and described. The figures are not necessarily drawn to scale, and in some instances the drawings have been exaggerated and/or simplified in places for illustrative purposes only. One of ordinary skill in the art will appreciate the many possible applications and variations of the present invention based on the following illustrative embodiments of the present invention.

[0014] FIGs. 2-11 show process steps for two illustrative embodiments of the present invention. An embodiment of the present invention provides a way to reduce defect density in an improved strained silicon structure, as compared to the prior design (see FIG. 1A), for example. One of the primary mechanisms for achieving reduced defect density in an embodiment of the present invention is providing free surfaces for grain boundaries so that dislocations will migrate to the free surfaces. By strategically locating such free surfaces away from the channel region (where the transistor will reside), dislocations may be moved away from the channel region (i.e., towards or to the free surfaces). The following description of these two illustrative embodiments will illustrate how an embodiment of the present invention may provide such free surfaces and thereby reduce defect density at the channel region.

[0015] FIGs. 2-7 show process steps for a first illustrative embodiment of the present invention. Beginning at FIG. 2, a first lattice mismatched epitaxial layer 40 is formed on a substrate 42. In this example, the substrate 42 is silicon (e.g., silicon wafer) and the first layer 40 is a SiGe buffer layer. The first layer 40 preferably has lattice mismatch relative to the substrate 42.

[0016] Next, as shown in FIG. 3, a first trench 44 is formed in the first layer 40. In FIG. 3, two parts of the first trench 44 are shown. The plan-view layout of the first trench 44 may vary, depending upon the layout of the active areas. Also, the depth and width of the first trench 44 may vary. In this case (FIG. 3), the first trench 44 extends through the first layer 40 and into the substrate 42. It is preferable that the first trench 44 extends through the interface 46 of the first layer 40 and the substrate 42. In other embodiments (not shown), the first trench 44 may extend only partially through the first layer 40 or only through the first layer 40 and not into the substrate 42.

[0017] The first layer 40 may be annealed after forming the first trench 44. Such annealing may be performed at an annealing temperature that is about 100° C higher than the deposition temperature used in forming the first layer 40, for example. Annealing the first layer 40 may remove dislocations or defects in the crystalline structure of the first layer 40, and/or may cause dislocations within the first layer 40 to migrate to or towards a free surface (e.g., at the first trench 44). The top surface 48 of the first layer 40 may be planarized (before or after the formation of the first trench 44) prior to forming the second layer 50 thereon (described further below). The planarization may be performed by any appropriate planarization process, such as chemical-mechanical polishing (CMP), for example. If the first layer 40 is annealed, it is preferable to perform the planarization after the annealing, but the planarization may be performed before the annealing. In other embodiments, the annealing of the first layer 40 and/or the planarizing of the first layer 40 may not be performed.

[0018] As shown in FIG. 4, a second lattice mismatched epitaxial layer 50 is formed on the first layer 40. In this example, the second layer 50 is a relaxed SiGe layer. As shown in FIG. 4, material of the second layer 50 may partially fill the first trench 44 (see portions 52 in FIG. 4).

Alternatively, the first trench 44 may be filled or partially filled with another material (e.g., insulating material) (not shown) prior to forming the second layer 50 on the first layer 40. Although the first trench 44 remains open after forming the second layer 50 in this example (see FIG. 4) and effectively extends up through the second layer 50, the first trench 44 may be closed by the second layer 50. This will depend on the overhangs 54 formed, if any, by the second layer material at the first trench 44 and the width of the first trench 44. It is preferable, but not necessary, that the first trench 44 remains open at this stage.

[0019] As with the first layer 40, the second layer 50 may be annealed. Such annealing may be performed at an annealing temperature that is about 100° C higher than the deposition temperature used in forming the second layer 50, for example. Annealing the second layer 50 may remove dislocations or defects in the crystalline structure of the second layer 50, and/or may cause dislocations within the second layer 50 to migrate to or towards a free surface (e.g., at the first trench 44). The top surface 56 of the second layer may be planarized prior to forming the third layer 60 thereon (described further below). If the second layer 50 is annealed, it is preferable to perform the planarization after the annealing, but the planarization may be performed before the annealing. In other embodiments, the annealing of the second layer 50 and/or the planarizing of the second layer 50 may not be performed.

[0020] In FIG. 5, a third lattice mismatched epitaxial layer 60 has been formed on the second layer 50. The third layer 60 is strained (under biaxial tension) due to the lattice mismatch between the second layer 50 and the third layer 60. The third layer 60 in this case is strained silicon under biaxial tension (see e.g., FIG. 1C). As shown in FIG. 5, the material of the third layer 60 may partially fill the first trench 44 (see portions 62 in FIG. 5). Although the first trench 44 remains open after forming the third layer 60 in this example (see FIG. 5), the first

trench 44 may be closed by the formation of the third layer 60. This will depend on the overhangs formed, if any, by the third layer material 60 at the first trench 44 and the width of the first trench 44.

[0021] The first and second layers 40, 50 may each have a thickness of about 2-3 μm , for example. The third layer 60 needs to be thin enough that it will not generate dislocations therein. If the third layer 60 is too thick, a crack or dislocation may form due to stress from the lattice constant difference between the second and third layers 50, 60. The third layer 60 may have a thickness of about 200 Å, for example. As mentioned above, the first and second layers 40, 50 may be SiGe layers. In such case, the second layer 50 preferably has a higher concentration of germanium than the first layer 40 to create a lattice mismatch between them. The first layer 40 may be a graded SiGe layer, for example. The first (40), second (50), and third (60) layers, each may be different and each may be formed from any of a variety of materials or combinations/compounds of materials, including (but not limited to): silicon, germanium, carbon, compound semiconductors, and combinations thereof, for example.

[0022] Although the first and second layers 40, 50 are each shown as single layers, either or both may be composite layers (i.e., a layer made from multiple layers). For example, the first layer 40 may include a graded composition layer and a uniform composition layer. Similarly, the second layer 50 may include a graded composition layer and a uniform composition layer, for example. Although the substrate 42 is shown in the illustrative embodiments herein as a silicon wafer, the substrate 42 may be a silicon-on-insulator (SOI) structure (not shown), or the substrate 42 may include an insulator layer over a silicon layer (not shown), for example. With the benefit of this disclosure one of ordinary skill in the art may realize many other possible material variations and/or layer compositions for an embodiment of the present invention.

[0023] In forming the first (40), second (50), and third (60) layers, any of a variety of epitaxial layer formation processes/technologies (or combinations thereof) may be used, including (but not limited to): chemical vapor deposition (CVD), MOCVD, H-CVD, atomic layer deposition, strained silicon molecular beam epitaxy (SS-MBE), and combinations thereof, for example.

[0024] In FIG. 6, a second trench 64 has been formed in the third and second layers 60, 50. At least part of the second trench 64 is in alignment (i.e., horizontal alignment) with at least part of the first trench 44, as shown in FIG. 6. In other words, the second trench 64 is located over the first trench 44. However, the second trench 64 need not be centered relative to the first trench 44 while being aligned with it. Preferably, the second trench 64 is aligned with the first trench 44 and has a depth so that the second trench 64 connects with and opens to the first trench 44. This will depend on the depth of the second trench 64 into the second layer 50 and whether the second layer material (i.e., overhang portions 54 of second layer 50 in first trench 44) closes the first trench 44 at the bottom of the second trench 64. In a preferred embodiment, the second trench 64 has a depth of about 300 Å, for example.

[0025] In the example embodiment shown in FIG. 6, the second trench 64 extends through the third layer 60 and partially through the second layer 50. In other embodiments (not shown), the second trench 64 may extend through the second layer 50 but not into the first layer 40. In still other embodiments (not shown), the second trench 64 may extend through the second layer 50 and into the first layer 40. The width of the second trench 64 is preferably wider than the first trench 44. However, in other embodiments (not shown), the width for part or all of the second trench 64 may be smaller than or equal to the width for part or all of the first trench 44,

for example. Preferably, the second trench 64 is the same as that normally used for shallow trench isolation (e.g., see STI 36 shown in FIG. 1A).

[0026] In FIG. 7, the second trench 64 and the unfilled portions of the first trench 44 have been filled with an insulating material 68. Hence, the filled second trench 64 acts as a isolation region for the active area. Also shown in FIG. 7, for purposes of illustration, a transistor 28 has been formed. Part of the transistor 28 is formed in the third layer 60 (i.e., in the strained silicon layer in this example).

[0027] The first trench 44 provides a free surface for dislocations to migrate to, which allows defects to be moved away from the channel region of the transistor 28. It is desirable to eliminate or reduce the number of defects in the channel region. Dislocations in the channel region may cause electrical leakage, for example. The first trench 44 is located under the shallow trench isolation (STI) region (i.e., the second trench 64). This provides the advantage of moving dislocations and/or defects to the region beneath or at the STI region. It is generally not problematic to have defects at or below the STI region, as it is away from the channel region and typically will not affect the performance of the device(s) formed in the active region. Also, the first trench 44 crosses the interface 72 of the first and second layers 40, 50. This relieves stress at this interface 72 to reduce or eliminate dislocations under the channel region.

[0028] FIGs. 8-11 show process steps for a second illustrative embodiment of the present invention. The second embodiment shown in FIGs. 8-11 is similar to the first embodiment shown in FIGs. 2-7, except that the first trench 44 is formed after the second layer 50 is formed (rather than before the second layer 50 is formed). Beginning at FIG. 8, a first lattice mismatched epitaxial layer 40 (e.g., SiGe layer) is formed on a substrate 42 (e.g., silicon wafer), and a second lattice mismatched epitaxial layer 50 (e.g., relaxed SiGe layer of higher Ge

concentration than the first layer 40) is formed on the first layer 40. The first layer 40 may be annealed and/or planarized before the second layer 50 is formed thereon (e.g., as described above with respect to the first embodiment).

[0029] As shown in FIG. 9, a first trench 44 is formed in the second and first layers 50, 40. In the example embodiment shown in FIG. 9, the first trench 44 extends through the second layer 50, through the first layer 40, and into the substrate 42, which is preferred. In other embodiments (not shown), the first trench 44 may extend through the second layer 50 and into the first layer 40 (but not into the substrate 42). In such cases, the first trench 44 may extend partially or entirely through the first layer 40. The second layer 50 may be annealed and/or planarized before the third layer 60 is formed thereon (e.g., as described above with respect to the first embodiment).

[0030] In FIG. 10, a third lattice mismatched epitaxial layer 60 (e.g., silicon) is formed on the second layer 50. The third layer 60 is strained due to the lattice mismatch between the second layer 50 and the third layer 60. The third layer material 62 may partially fill the first trench, as shown in FIG. 10. In FIG. 11, a second trench 64 has been formed in alignment with the first trench 44 (as described above regarding the first embodiment), and the second trench 64 has been filled with an insulating material 68 to form an isolation region. The insulating material 68 also fills the open remainder of the first trench 44, as shown in FIG. 11. The second trench 64 may extend partially or completely through the second layer 50. In another embodiment (not shown), the second trench 64 may extend to, partially into, and/or completely through the first layer 40, for example. As in FIG. 7, a transistor 28 is shown in FIG. 11 for purposes of illustrating an example use of a strained silicon structure.

[0031] In yet another embodiment of the present invention (not shown), e.g., as a variation upon and/or additional step to the first embodiment and/or the second embodiment, a trench may be formed in the substrate 42 before the first layer 40 is formed on the substrate 42. Depending upon the depth and width of such a trench in the substrate 42, it may or may not be filled by the material of the first layer 40 when the first layer is formed. Preferably, such a trench in the substrate 42 is depth enough and/or wide enough that the trench extends through the first layer 40 after the first layer is formed (in spite of overhang and filling by the first layer material), as the second layer 50 does in FIG. 4, for example. With the benefit of this disclosure, one of ordinary skill in the art may realize many other variations and embodiments of the present invention.

[0032] It will be appreciated by those skilled in the art having the benefit of this disclosure that embodiments the present invention provide methods of forming a strained silicon structure. It should be understood that the drawings and detailed description herein are to be regarded in an illustrative rather than a restrictive manner, and are not intended to limit the invention to the particular forms and examples disclosed. On the contrary, the invention includes any further modifications, changes, rearrangements, substitutions, alternatives, design choices, and embodiments apparent to those of ordinary skill in the art, without departing from the spirit and scope of this invention, as defined by the following claims. Thus, it is intended that the following claims be interpreted to embrace all such further modifications, changes, rearrangements, substitutions, alternatives, design choices, and embodiments.